

Міністерство освіти і науки України
Національний університет водного господарства та
природокористування
Навчально-науковий інститут автоматики, кібернетики і
обчислювальної техніки
Кафедра вищої математики

04-02-46М

МЕТОДИЧНІ ВКАЗІВКИ
до виконання практичних робіт з навчальної дисципліни
**«МАТЕМАТИЧНА СТАТИСТИКА. МАТЕМАТИЧНІ
МЕТОДИ В ПСИХОЛОГІЇ»**
для здобувачів вищої освіти першого (бакалаврського)
рівня за освітньо-професійною програмою «Психологія»
спеціальності 053 «Психологія»
денної та заочної форм навчання

Рекомендовано
науково-методичною
радою з якості ННІ АКОТ
Протокол № 3
від 29.12.2020 р.

Рівне – 2020

Методичні вказівки до виконання практичних робіт з навчальної дисципліни «Математична статистика. Математичні методи в психології» для здобувачів вищої освіти першого (бакалаврського) рівня за освітньо-професійною програмою «Психологія» спеціальності 053 «Психологія» денної та заочної форм навчання. [Електронне видання] / Кушнір О. О., Кушнір В. П. – Рівне : НУВГП, 2020. – 76 с.

Укладачі:

Кушнір О. О., канд. фіз.-мат. наук, доцент кафедри вищої математики;

Кушнір В. П., канд. фіз.-мат. наук, доцент кафедри вищої математики.

Відповідальний за випуск: Тадеєв П. О., доктор пед. наук, проф., зав. кафедри вищої математики.

Керівник групи забезпечення спеціальності «Психологія»

Шпак С. Г.,
к.пс.н., доцент.

© Кушнір О. О.,
Кушнір В. П., 2020
© НУВГП, 2020

Зміст

Вступ.....	4
Встановлення програми PSPP.....	4
§ 1. Упорядкування даних.....	6
§ 2. Оцінка числових характеристик.....	14
§ 3. Перевірка гіпотези про розподіл.....	22
§ 4. Перевірка наявності зсуву значень.....	28
§ 5. Виявлення відмінностей у рівнях.....	37
§ 6. Перевірка однорідності двох вибірок.....	43
Таблиця критичних значень однобічного критерію серій Вальда-Вольфовиця рівня значущості 0,05 ($n \geq m$).....	45
§ 7. Виявлення різниці між імовірностями.....	53
§ 8. Кореляційний аналіз.....	58
§ 9. Дисперсійний аналіз.....	67
§ 10 Регресійний аналіз.....	72
Література.....	76

Вступ

Для обробки даних у психології, як і в інших науках, на зміну ручним обчисленням прийшли спеціальні статистичні програми, які дозволяють швидко і точно виконувати необхідні процедури.

Однією з найбільш доступних статистичних програм, якою зручно користуватися психологам, є PSPP. Її можна безплатно встановити на домашній комп'ютер чи ноутбук з довільною операційною системою. Ця програма є сумісною з іншою популярною серед психологів статистичною програмою SPSS: вони працюють з файлами однакового типу. Результати можна зберігати у поширених форматах з розширеннями .html, .txt, .odt, .pdf, .ps, .csv.

З метою надання студентам методичної допомоги до кожної теми з дисципліни "Математична статистика. Математичні методи в психології" подано короткі теоретичні відомості та наведені алгоритми виконання типових завдань за допомогою статистичної програми PSPP.

Встановлення програми PSPP.

1. В Googl набираємо тільки 4 букви: PSPP.
2. Вибираємо [PSPP - GNU Project - Free Software Foundation - Gnu.org](#)

3. В розділі Downloading PSPP вибираємо [download or otherwise obtain](#)
4. В розділі Binaries в пункті Windows: (якщо у Вас встановлена саме ця операційна система) вибираємо [all downloadable files](#)
5. Якщо у Вас 32-бітна версія Windows, то двічі клацаємо мишкою по зеленому полю (pspp-20200905-...).
6. Чекаємо, поки завантажаться файли, а потім натискаємо в лівому нижньому кутку на стрілочку і починаємо встановлювати програму на свій комп'ютер.
7. Якщо у Вас 32-бітна версія Windows, то замість усіх вказаних вище пунктів завантажите та встановіть 32-бітну версію PSPP, перейшовши за посиланням

<https://sourceforge.net/projects/pspp4windows/files/2020-09-05/pspp-20200905-daily-32bits-setup.exe/download>

Якщо станеться так, що програма завантажилася англійською мовою, то **мову можна змінити** наступним способом:

1. Натискаєте на ярлик правою кнопкою миші.
2. Знизу натискаєте Свойства.
3. Заходите на вкладку Ярлык.
4. У полі Объект вносите такі виправлення:

а)на початку вставляєте C:\Windows\System32\cmd.exe /c "SET LANG=uk_UA && START /D ^"

б)після bin замість \ перед psppire.exe" вставляєте ^" та пропуск.

Можете також в полі Окно замінити Обычный размер окна на Свернутое в значок. (Для того, щоб не з'являлося на секунду чорне вікно перед запуском програми.)

(Інструкція взята з сайту <https://stackoverflow.com/a/>)

Після цього програма буде українською мовою.

Приклад. В полі Обьект було

"C:\Program Files\PSPP\bin\psppire.exe"

Замінили на C:\Windows\System32\cmd.exe /c "SET LANG=uk_UA && START /D ^"C:\Program Files\PSPP\bin\psppire.exe"

Завантажилася PSPP українською.

§ 1. Упорядкування даних

Мета роботи: Навчитися вводити дані та впорядковувати їх у програмі PSPP, познайомитися з поняттям квантилів.

Вступна інформація. Введення даних — це найбільш трудомістка справа при використанні PSPP. Звичайно, мова не йде про використання вже готових даних в електронній формі, наприклад, у форматі MS EXEL чи таблиць у MS WORD. Ці дані уже були кимось

внесені раніше. Але нові дані, які щойно були отримані, потрібно ввести, щоб потім їх можна було обробляти та аналізувати.

Якщо при введенні даних була допущена помилка, її можна легко виправити. Для цього потрібно перевести курсор у відповідну комірку, внести зміни та натиснути ENTER. Якщо потрібно видалити змінну повністю, виділяємо відповідний стовпчик і натискаємо Спорожнити змінну.

Потім дані в програмі PSPP уже легко подавати в інших формах: з вибірки утворювати варіаційний ряд, дискретний варіаційний ряд, інтервальний варіаційний ряд, знаходити емпіричну функцію розподілу. Також у програмі можна швидко знаходити процентилі, міри центральної тенденції, міри мінливості та інші числові характеристики.

Назви змінних мають починатися з букви, а не цифри та містити не більше 8 символів. У деяких версіях програми PSPP не зберігаються українські літери.

Шкала вимірювання змінних задається у стовпчику Міра. Числова шкала називається Масштаб, а шкала назв — Іменна. Для дихотомічних змінних спеціальна назва не передбачена. Тому їх можна віднести до шкали Порядкова, а значення надавати тільки 1 і 0.

Першою змінною варто вводити ініціали респондента. Повністю прізвище писати не варто. Над цією змінною жодних дій не будемо виконувати, тому її

Тип буде Рядок, а Міра — Іменна. Усім іншим змінним варто приписувати Тип — Число.

Завдання роботи № 1:

1. Ввести дані в програму PSPP.
2. Утворити варіаційний ряд змінної Середній бал атестата.
3. Знайти правобічну границю в кожній точці емпіричної функції розподілу (або накопичену відносну частоту) змінної Середній бал атестата.
4. Знайти вибіркві квінтилі та кватилі цієї змінної.
5. Знайти міри центральної тенденції: вибіркві моду, медіану, середнє значення та міри мінливості: розмах, виправлену дисперсію та оцінку середнього квадратичного відхилення.
6. Для альтернативних змінних побудувати кругові та стовпчикові діаграми, а для числових — гістограми.
7. Зробити висновки.

Порядок виконання роботи № 1

1. Створити теку під своїм прізвищем англійською та в ній підтеку ...1 (замість трикрапки своє прізвище англійською).
2. Завантажити програму PSPP.
3. У вікні “Перегляд змінних” вводимо назви змінних.

Можна ввести наступні 8 назв: н, ін, в, к, укр, п2, п3, сб.

4. Зберегти файл у теці ...1 під такою ж назвою ...1.sav у форматі “системний файл”. Дані в PSPP автоматично не зберігаються, тому **рекомендується їх часто зберігати.**
5. У вікні “Перегляд змінних” вводимо атрибути змінних
типи: другий — рядок, усі інші — число;
ширина (кожного разу натискаємо enter двічі) — 2, 3, 1, 1, 3, 3, 3, 4 (зберегти файл);
Знаків після коми — 0, 0, 0, 0, 0, 0, 0, 1 (зберегти файл);
мітки: (краще набирати з великої літери) 1 — Порядковий номер, 2 — Ініціали, 3 — Вибір 3-го предмета, 4 — Сільський коефіцієнт (якщо у файлі з даними не сказано інакше), 5 — Українська мова та література (ЗНО), 6 — (вказати назву 2-го предмета, яка написана у файлі даних під Українська...), 7 — Бали ЗНО з 3-го предмета, 8 — Середній бал атестата (зберегти файл);
стовпчики — 2, 3, 1, 1, 2, 2, 2, 3 (зберегти файл);
міра: 2 — Іменна, 3, 4 — Порядкова, решта — Масштаб (зберегти файл).
6. Вказати мітки значень 2-ї змінної, натиснувши на клітинку та заповнивши табличку: 0 — ... (назва

предмета, вказана у файлі даних вгорі), натискаємо Додати, 1 — ... (назва предмета там же) натискаємо Додати і Гаразд (зберегти файл); вказати мітки значень 3-ї змінної: 0 — немає сільського (чи галузевого) коефіцієнта, Додати, 1 — є сільський (чи галузевий) коефіцієнт, Додати і Гаразд (зберегти файл).

7. У вікні “Перегляд даних” вводимо дані, **регулярно зберігаючи файл**. Із прізвища беремо тільки 1-шу літеру та ініціали, записуємо їх латиницею. Кожного разу натискаємо Таб. Десятковий знак відділяється комою.
8. Розпочинаємо вивчення змінної сб. Маємо вибірку. Перетворимо її у варіаційний ряд. Для цього вибираємо в меню Дані — упорядкувати спостереження. Вибираємо критерій: Середній бал атестата і переносимо його за допомогою стрілочки. Оскільки деякі значення повторюються, то можна вибрати ще один критерій: Українська... Упорядковуємо за зростанням. Натискаємо Гаразд.
9. Отримали варіаційний ряд змінної сб. За ним відразу можемо відшукати вибірккову медіану — значення, яке знаходиться посередині. Зробіть це.
10. Створюємо текстовий файл під своїм прізвищем (українською) у теці Робота1 та записуємо в ньому

знайдене значення вибіркової медіани.

11. Знайдемо емпіричну функцію розподілу (точніше її правобічну границю у кожній точці або накопичені відносні частоти). Для цього вибираємо в меню Перетворення — Ранжувати спостереження. Вибираємо змінну Середній бал атестата і переносимо її у поле Змінні. Поле Критерій залишаємо порожнім. Типи рангів — Дробовий ранг. Продовжити. Збіги — Найбільший. Продовжити. Натискаємо Гаразд. Отримали 8-му змінну — Rсб. Зберігаємо файл.
12. Оскільки ми хочемо бачити проценти цілими, переходимо до вікна Перегляд змінних. Для 8-ї змінної встановлюємо Ширина — 3 (натискаємо enter), Знаків після коми — 2, Столпчики — 2 (натискаємо enter). Знову зберігаємо файл.
13. Знаходимо вибіркові квінтилі, тобто значення змінної сб, при яких правобічна границя емпіричної функції розподілу (змінна Rсб) досягне або перевищить 0,2, 0,4, 0,6 та 0,8 і вибіркові квінтилі — 0,25, 0,5, 0,75. Якщо обсяг вибірки — непарний, то середній квінтиль є вибірковою медіаною. Переконайтеся в цьому. Записуємо знайдені значення у свій текстовий файл.
14. Таким чином, підготовча робота проведена. У файлі

“Вивід PSPP” наводиться інформація про неодноразове збереження файлу. Оскільки ця інформація не буде потрібною, варто файл “Вивід PSPP” закрити.

15. Побудуємо кругові діаграми для альтернативних змінних. У меню вибираємо Аналіз — Описова статистика — Частоти. Переносимо змінні Сільський коефіцієнт та Вибір 3-го предмета. Знімаємо виділення усіх полів Статистики. Натискаємо кнопку Діаграми. Виділяємо поля Малювати кругові діаграми та Малювати стовпчикові діаграми. Продовжити. Натискаємо Гаразд. У файлі “Виведення PSPP” отримуємо таблицю частот та діаграми. Тепер файл “Вивід PSPP” експортуємо у файл ...1.pdf у теці Робота1. Для цього вибираємо в меню Файл—Експортувати. Набираємо назву — ...1 (замість трикрапки своє прізвище англійською). Знизу вибираємо Тип файла: pdf. Зліва вибираємо теку ...1. Натискаємо Зберегти. Після цього закриваємо “Вивід PSPP”.
16. За діаграмами візуально оцінюємо частку абітурієнтів, які мають сільський коефіцієнт та який 3-й предмет абітурієнти вибирали частіше (вибіркові моди). Записуємо свої висновки у текстовому файлі.
17. Знайдемо деякі статистики за допомогою PSPP. У

меню вибираємо Аналіз — Описова статистика — Частоти. Переносимо змінну Середній бал атестата (сб). Вибираємо статистики Середнє, Стандартне відхилення, Мінімум, Максимум, Розсіювання, Діапазон, Режим, Медіана. Натискаємо кнопку Діаграми. Виділяємо поля Малювати гістограми та Накласти криву нормального розподілу. Продовжити. Натискаємо Гаразд.

18. У меню вибираємо Аналіз — Описова статистика — Частоти. Вилучаємо стрілкою сб та переносимо три інші числові змінні — бали ЗНО з української та з двох інших предметів. Натискаємо Гаразд. У файлі “Вивід PSPP” отримуємо таблиці частот, обчислені значення статистик та гістограми. Тепер файл “Вивід PSPP” експортуємо у файл ...11.pdf (замість трикрапки вводимо своє прізвище англійською) у теці ...1.

19. Із таблиці файлу ...11.pdf виписуємо Обсяг вибірки (N (Чинний)); Міри центральної тенденції: Середнє; Медіана, та Мода (Режим). (Якщо найбільшу частоту мають декілька несусідніх значень змінної, то вибіркова мода не існує.) Міри мінливості: Розмах вибірки R= (Діапазон) — різниця між максимумом та мінімумом, Нормована оцінка дисперсії (Розсіювання),

Стандартне відхилення.

20. Візуально оцінюємо за гістограмами близькість вибірових розподілів до нормальних. Записуємо свої висновки в текстовому файлі та зберігаємо його.
21. Надсилаємо усі файли теки ...1 на електронну пошту викладачу для перевірки.

§ 2. Оцінка числових характеристик

Мета роботи: Навчитися переводити “сирі” оцінки в стандартну шкалу стенов, оцінювати асиметрію та ексцес, знаходити інтервальні оцінки параметрів нормального розподілу та ймовірності значень ознаки, навчитися знаходити квантілі розподілу χ^2 Хельмерта-Пірсона.

Вступна інформація. Для того щоб дані щодо однієї й тієї ж характеристики, отримані за різними методиками, можна було порівнювати, їх переводять у стандартні шкали: z-оцінки, Т Мак-Колла, IQ Векслера, стени Кеттеля, стенови, 11-бальна шкала.

В програмі PSPP є стандартні процедури для знаходження надійних інтервалів для середнього значення генеральної сукупності.

Також передбачена можливість використання вбудованих функцій. Серед них є математичні, зокрема $\text{SQRT}(x)$ — квадратний корінь з числа x , а також

статистичні.

Із CDF починаються назви стандартних функцій розподілу. У дужках спочатку вказується аргумент, а потім — параметри розподілу.

З IDF починаються назви функцій, обернених до стандартних функцій розподілу. З їх допомогою можна знаходити квантілі. Щоб знайти квантиль рівня α , в дужках спочатку вказується $1-\alpha$, а потім — параметри розподілу.

Із PDF починаються щільності стандартних розподілів.

Розподіл χ^2 Хельмерта-Пірсона позначається CHISQ, t-Ст'юдента — T, нормальний — NORMAL.

Для побудови двобічних надійних інтервалів надійності γ використовуються квантілі рівнів $(1+\gamma)/2$ та $(1-\gamma)/2$. Якщо $\gamma=0,95$, то потрібні квантілі рівнів 0,975 та 0,025.

Для побудови надійних інтервалів для ймовірності з використанням комп'ютерів можна використовувати безпосередньо біномний розподіл, а не його нормальне наближення.

Завдання роботи № 2:

1. Перевести оцінки “Середній бал атестата” в стандартну шкалу стенонів.

2. Знайти міжквартильне відхилення, середнє відхилення та квартильний коефіцієнт асиметрії.
3. Оцінити асиметрію та ексцес і зробити висновок про узгодженість даних із нормальним законом розподілу.
4. Побудувати інтервали надійності 0,95 та 0,99 для параметрів нормального розподілу за даними вибірки середнього балу атестата.
5. Побудувати інтервали надійності 0,95 для ймовірностей того, що абітурієнт проходив ЗНО по першому предметові та що в нього є сільський коефіцієнт.

Порядок виконання роботи № 2

1. Створити у своїй теці підтеку під назвою ...2 (замість трикрапки своє прізвище англійською).
2. Відкрити системний файл із попередньої практичної роботи та зберегти його у теці ...2 під назвою ...2 (замість трикрапки своє прізвище англійською).
3. Переведемо наші “сирі” бали у стандартну шкалу стенонів. Для цього вибираємо в меню Аналіз — Описова статистика — Описова статистика. Вибираємо змінну Середній бал атестата і переносимо її у вікно Змінні. Відмічаємо параметри — Зберегти Z-оцінки ... Натискаємо Гаразд.

4. Отримали нову змінну Zсб. Переходимо в Перегляд змінних і зменшуємо їй кількість стовпчиків до 3. Натискаємо enter. Зберігаємо системний файл.
5. Тепер переведемо Z-оцінки в стенайни. Для цього оголошуємо нову змінну Ссб. Ширина — 2, Знаків після коми — 0, Мітка — Стенайни середнього бала атестата, Стовпчики — 1. Зберігаємо системний файл.
6. В меню вибираємо Перетворення — Обчислити. Набираємо вручну змінну для обробки Ссб. У віконечку Числові вирази набираємо 5+2 (знак множення). Вибираємо змінну Z-оцінка... і переносимо її за допомогою стрілочки. Натискаємо Гаразд. Зберігаємо системний файл. Отримали оцінки у шкалі стенайнів, точніше: не в 9-бальній, а в 11-бальній шкалі, оскільки можливі оцінки від 0 до 10.
7. Продовжуємо вивчати змінну сб. Щоб очистити файл Вивід, закриємо його.
8. Вибираємо в меню Аналіз — Описова статистика — Дослідити. Переносимо змінну Середній бал атестата у поле Список залежних.
9. Натискаємо Статистика. Відмічаємо усі 3 поля. Продовжити. Натискаємо Гаразд.
10. Експортуємо файл Вивід у файл із

розширенням ...2.odt у теці ...2.

11. Тепер будемо аналізувати інформацію у файлі ...2.odt. У першій табличці під літерою N записане число, яке є обсягом вибірки. Напишемо під таблицею “Обсяг вибірки N=...”.
12. Під другою табличкою напишемо “Найвищий середній бал атестата ... в абітурієнта ... (подивитися в 2-му стовпчику даних в останньому рядку), а найнижчий ... в абітурієнта ...” (в 1-му рядку).
13. Під 3-ю табличкою пишемо: Вибіркові квартилі $Q_1=...$, $Q_2=...$, $Q_3=...$ (вказуємо Кутові точки Тукі). Вибіркова медіана $Md=...$ (вказуємо Q_2). Оцінки квартилів генеральної сукупності $Q'_1=...$, $Q'_2=...$, $Q'_3=...$ (вказуємо зважені середні процентилів 25, 50, 75). Зауважимо, що вибіркова медіана дорівнює оцінці медіани генеральної сукупності.
14. Під 4-ю табличкою записуємо міри мінливості: нормовану оцінку дисперсії — випишуємо з таблиці значення Розсіювання; серединне відхилення E — половина міжквартильного відхилення (у таблиці вказана різниця між оцінками квартилів генеральної сукупності).
15. Квартильний коефіцієнт асиметрії обчислюємо за формулою $As(Q) = (Q_3 + Q_1 - 2Md) / (Q_3 - Q_1)$. Він, а також статистики Асиметрія та Ексцес, які вже

обчислила за нашою вказівкою програма PSPP, показують, наскільки сильно вибірковий розподіл відхиляється від нормального. Абсолютне значення Асиметрії та Ексцесу має бути меншим від потроєної відповідної Стандартної похибки. Перевірте це і напишіть у файлі ...2.odt. Якщо ці нерівності виконуються, то напишіть: “Немає підстав відкидати гіпотезу про нормальний розподіл”, а якщо — ні, то “Гіпотезу про нормальний розподіл слід відхилити”.

16. Припустивши, що розподіл змінної s_b — нормальний, побудуємо надійні інтервали для його параметрів. Надійний інтервал для математичного сподівання з надійністю 0,95 виписуємо з 4-ї таблиці (від Нижня межа до Верхня межа). Зберігаємо файл ...2.odt.
17. Для знаходження надійного інтервалу надійності 0,99 знаходимо в меню Аналіз — Порівняти середні — Одновибіркова t-перевірка. Переносимо змінну Середній бал атестата, а Значення для перевірки вводим 0. Натискаємо Параметри та змінюємо Довірчий інтервал із 95% на 99%. Натискаємо Продовжити, Гаразд.
18. Експортуємо файл Вивід у ...21.odt. Скопіюємо його вміст і перенесемо в файл ...2.odt.

19. Випишуємо в файлі ...2.odt надійний інтервал для математичного сподівання з надійністю 0,99. Зберігаємо файл ...2.odt.
20. Для знаходження надійного інтервалу для дисперсії знаходимо в меню Перетворення — Обчислити. Змінна для обробки набираємо Межа1. У поле Числові вирази вводим значення виправленої дисперсії (виписаного в файлі ...2.odt під 4-ю таблицею, в таблиці — Розсіювання) з крапкою, а не комою; знак множення; кількість ступенів вільності (обсяг вибірки, зменшений на 1); знак ділення.
21. Далі вибираємо функцію IDFCNISQ... та переносимо її стрілочкою догори. Замість першого знака питання вводим 0.975 (з крапкою, а не комою), а замість другого — кількість ступенів вільності. Натискаємо Гаразд.
22. У вікні Перегляд даних з'явилася ліва межа надійного інтервалу для дисперсії надійності 0,95. Зберігаємо системний файл.
23. Тепер знову Перетворення — Обчислити. Змінна для обробки набираємо Межа2. Цього разу замість 0.975 вводим 0.025. Натискаємо Гаразд. Зберігаємо системний файл.
24. У файлі ...2.odt записуємо надійний інтервал для дисперсії надійності 0,95. Зберігаємо файл ...2.odt.

25. Аналогічно знаходимо інтервал надійності 0,99 для дисперсії та середнього квадратичного відхилення. Цього разу використаємо змінні Межа3 та Межа4 і набиратимемо 0.995 та 0.005. Записуємо знайдені інтервали у файл ...2.odt та зберігаємо його.
26. Для знаходження меж надійного інтервалу для середнього квадратичного відхилення добуваємо квадратний корінь із меж для дисперсії. Записуємо обчислення у файл ...2.odt та зберігаємо його.
27. Тепер перейдемо до аналізу дихотомічних змінних в та ск. Оцінимо ймовірності того, що абітурієнт проходив ЗНО по першому предметові та що в нього є сільський коефіцієнт із надійністю 0,95. Для цього в меню вибираємо Аналіз — Описова статистика — Частоти. Знімаємо виділення усіх полів Статистики. Натискаємо кнопку Діаграми. Знімаємо виділення з Малювати гістограми. Переносимо стрілочкою змінні Вибір... та Сільський коефіцієнт. Натискаємо Гаразд.
28. У файлі Вивід отримуємо шукані оцінки ймовірностей у процентах у таблицях частот. Експортуємо цей файл у ...22.odt. Скопіюємо його вміст і перенесемо в файл ...2.odt. Закриваємо файл Вивід.
29. Будуємо надійні інтервали для ймовірностей подій.

Для цього знаходимо в меню Аналіз — Непараметричні критерії — Біноміальний. Переносимо Вибір... та Сільський коефіцієнт у поле Список змінних для перевірки. У полі Частка, що перевіряється підбираємо число 0... (з двома знаками після крапки) спочатку менше від спостережуваної частоти, таке, щоб у файлі Вивід Точна знач. (1-бічна) стала 0,025 або трохи меншою (Якщо 2-бічна, то 0,05). Натискаємо Гаразд. Зберігаємо системний файл. Саме це число й буде нижньою межею шуканого надійного інтервалу для відповідної ймовірності. Потім так само підбираємо верхню межу, яка є більшою від спостережуваної частоти. Зберігаємо системний файл. Експортуємо файл Вивід у ...23.odt. Копіюємо в ньому інформацію і переносимо у файл ...2.odt.

30. Записуємо в ...2.odt знайдені надійні інтервали для ймовірностей. Зберігаємо файл.
31. Надсилаємо усі файли теки ...2 на електронну пошту викладачу для перевірки.

§ 3. Перевірка гіпотези про розподіл

Мета роботи: Навчитися перевіряти гіпотезу про нормальний розподіл за критеріями λ -Колмогорова та χ^2 -Пірсона.

Вступна інформація. У програмі PSPP дуже легко

перевіряти гіпотезу про розподіл за допомогою класичного λ -критерію Колмогорова (без поправок та заміни критичних точок для складних гіпотез).

Щоб перевірити гіпотезу про розподіл за χ^2 -критерієм Пірсона, потрібно дані розбити на групи. Це можна зробити різними способами, і від цього залежатиме потужність критерію. Як правило, потужність буде більшою, якщо групи даних вибирати так, щоб були рівними не довжини проміжків, а теоретичні частоти. Для визначення кількості груп, залежно від обсягу вибірки, існує багато різних формул, але користування ними не гарантує максимальної потужності. У дослідженні (https://ami.nstu.ru/~headrd/seminar/publik_html/Z_lab_8.htm) вказується, що часто для досягнення максимальної потужності найкраще брати мінімально можливу для нормального розподілу кількість проміжків — 4 (бо кількість ступенів вільності обчислюється за формулою $f=k-3$, де k — кількість проміжків, оскільки 2 параметри нормального розподілу — математичне сподівання та дисперсія оцінюються за вибіркою). При цьому частоти проміжків будуть максимальними, тому сама можливість застосування критерію також зростає.

Принаймні кількість проміжків групування не може бути більшою від $N/5$, де N — обсяг вибірки (бо теоретичні частоти в проміжках мають бути не меншими від 5). Таким чином, для застосування χ^2 -критерію Пірсона необхідно, щоб обсяг вибірки був не менший від 20.

Завдання роботи № 3:

1. Перевірити гіпотезу про нормальний розподіл змінних “Середній бал атестата” та усіх 3-х результатів ЗНО за допомогою λ -критерію Колмогорова, використовуючи таблицю критичних значень Ліліфорса, та χ^2 -критерію Пірсона.
2. Зробити загальні висновки щодо розподілів та порівняння потужності різних критеріїв.

Порядок виконання роботи № 3

1. Створити у своїй теці підтеку під назвою ...3 (замість трикрапки своє прізвище англійською).
2. Відкрити системний файл із попередньої практичної роботи та зберегти його у теці ...3 під назвою ...3 (замість трикрапки своє прізвище англійською).
3. Залишаємо тільки перші 11 змінних. Решту (Межа1, ...) можна видалити. Для цього переходимо у вікно Перегляд даних, виділяємо непотрібний стовпчик, клацаємо правою кнопкою миші та натискаємо Спорожнити змінні. Зберігаємо системний файл ...3.sav. Закриваємо файл Вивід.
4. Перевіримо гіпотезу про нормальний розподіл змінних за допомогою λ -критерію Колмогорова. Для цього вибираємо в меню Аналіз — Непараметричні критерії — Одновибірковий Колмогорова-Смірнова.

Переносимо “Середній бал атестата” та усі три результати ЗНО за допомогою стрілочки. Вибираємо розподіл — Нормальний. Натискаємо Гаразд. Експортуємо файл Вивід у файл ...3.odt (замість трикрапки своє прізвище англійською) у теці ...3. Закриваємо файл Вивід.

5. У отриманій таблиці файлу ...3.odt знаходимо “Колмогорова-Смірнова (Z)”. Якщо ця величина більша від 1,035, то гіпотезу про нормальний розподіл слід відкинути на рівні значущості 0,01. Якщо ж вона менша від 0,895 (що відповідає рівневі значущості 0,05), то гіпотезу про нормальний розподіл за λ -критерієм Колмогорова немає підстав відкидати. Якщо вона більша від 0,895, але менша від 1,035, то гіпотеза про нормальний розподіл відкидається за критерієм Колмогорова на рівні значущості 0,05. Пишемо свої висновки у файлі ...3.odt окремо за кожною з 4-х змінних та обґрунтовуємо їх. Зберігаємо файл ...3.odt
6. Тепер перевіримо гіпотезу про нормальний розподіл змінної “Середній бал атестата” за χ^2 -критерієм Пірсона. Для цього (при $N < 100$) згрупуємо дані на 4 групи за допомогою квантилів z-розподілу:
 $Q_1 = -0,84$, $Q_2 = 0$, $Q_3 = 0,84$.
7. Вибираємо в меню Перетворення — Перекодувати

до інших змінних. Переносимо змінну “Z-оцінка...” за допомогою стрілочки. Натискаємо на Zсб і активізується вікно Нова змінна. У ній набираємо “сб4” (мітка — Середній бал атестата — 4 групи) і натискаємо Зміна.

8. Натискаємо Попередні і нові значення. Натискаємо кнопку “Діапазон значень від найменшого до вказаного” і вводимо —0,84, а в полі Значення вводимо 1. Натискаємо Додати. Далі натискаємо кнопку “Діапазон від” та вводимо значення —0,84 та 0 відповідно. У полі Значення вводимо 2. Натискаємо Додати. Далі вводимо значення 0 та 0,84 відповідно. У полі Значення вводимо 3. Натискаємо Додати. Далі натискаємо кнопку Усі інші значення, а в полі Значення вводимо 4. Знову натискаємо Додати. Продовжити — Гаразд.
9. Для змінної сб4 відводимо ширину 1 і 1 стовпчик; Міра — Порядкова. Зберігаємо ...3.sav.
10. Натискаємо Аналіз — Непараметричні критерії — χ -квадрат Переносимо змінну Середній бал атестата — 4 групи. Натискаємо Гаразд і експортуємо файл Вивід у ...31.odt. Закриваємо файл Вивід.
11. Переносимо інформацію із ... 31.odt у ...3.odt.
12. Звертаємо увагу, що в останній таблиці $df=3$, тобто

кількість ступенів вільності вказана неправильно, бо PSPP не знала, що ми два параметри — математичне сподівання та дисперсію — оцінювали за вибіркою. Тому й Асимпт. знач. вказане неправильно. Тому з цієї таблиці беремо тільки величину χ -квадрат і порівнюємо її з квантилем χ -квадрат -розподілу з 1 ступенем вільності рівня 0,95, який дорівнює 3,841, та рівня 0,99, який дорівнює 6,635. Якщо статистика χ -квадрат менша від 3,841, то записуємо у файлі ...3.odt: “За результатами критерію χ -квадрат немає підстав відкидати гіпотезу про нормальний розподіл на рівні значущості 0,05.” Якщо статистика χ -квадрат більша від 6,635, то гіпотеза про нормальний розподіл відкидається за критерієм Пірсона на рівні значущості 0,01. Якщо статистика χ -квадрат більша від 3,841, але менша від 6,635, то гіпотеза про нормальний розподіл відкидається за критерієм Пірсона на рівні значущості 0,05.

13. Аналогічно перевірте гіпотезу про нормальний розподіл балів ЗНО з 3-х предметів за χ -квадрат критерієм Пірсона. Для знаходимо їх z-оцінки (Аналіз — Описова статистика — Описова статистика), потім утворюємо нові змінні укр4, п24, п34 (діємо так само, як і випадку сб4), далі перевіряємо гіпотезу про нормальні розподіли за критерієм χ -квадрат, експортуємо Вивід у ...33.odt.

Закриваємо файл Вивід. Переносимо інформацію із ... 33.odt у ...3.odt. Робимо висновки.

14. Порівняйте результати перевірки гіпотези про нормальний розподіл за різними критеріями та зробіть загальні висновки щодо розподілів та порівняння потужності різних критеріїв, запишіть їх у файлі ...3.odt та збережіть його. Гіпотеза відкидається, якщо вона відкидається хоча б за одним критерієм. Рівень значущості вибирається мінімальний. Потужнішим є той критерій, за яким гіпотеза відкидається. Якщо гіпотеза відкидається за обома критеріями, то потужніший той, за яким гіпотеза відкидається при меншому рівні значущості.
15. Надішліть усі файли теки ...3 на електронну пошту викладачу для перевірки.

§ 4. Перевірка наявності зсуву значень.

Мета роботи: Навчитися перевіряти наявність зсуву значень зв'язаних змінних; робити обґрунтовані висновки.

Вступна інформація. Для перевірки наявності зсуву значень в програмі PSPP можна використовувати t -критерій Стюдента для парних вибірок, дисперсійний аналіз (ANOVA), χ^2 -критерій Фрідмана, T -критерій Вілкоксона та G -критерій знаків.

χ^2 -критерій Фрідмана можна використовувати в PSPP тільки при великих обсягах вибірки.

t -критерій Стюдента для парних вибірок можна використовувати не тільки тоді, коли їх обсяг великий, але й у випадку малих вибірок, якщо відомо, що вона взята з нормального розподілу.

Т-критерій Вілкоксона можна застосовувати в PSPP уже при обсягах вибірки $n > 14$.

G-критерій знаків можна використовувати як при великих, так і при малих обсягах вибірки. Розподіл змінної може бути довільним.

Для перевірки наявності зсуву значень альтернативних змінних в програмі PSPP є також Q-критерій Кохрена та критерій МакНемара. Але вони видають такі ж результати, як і вже згадані χ^2 -критерій Фрідмана та G-критерій знаків відповідно. По-суті, критерій МакНемара і є G-критерієм знаків у випадку альтернативних змінних.

Завдання роботи № 4:

1. Перевірити за χ^2 -критерієм Фрідмана, чи більша кількість абітурієнтів отримують значно кращі бали ЗНО з одного предмета, ніж із іншого.
2. Виявити, між якими саме предметами є ця суттєва різниця за критерієм Т Вілкоксона.
3. Виявити, чи з якогось предмета значно більша

кількість абітурієнтів отримують вищі бали ЗНО, ніж із іншого за G критерієм знаків.

4. Порівняти попарно середні результати ЗНО за критерієм t Стюдента для парних вибірок.
5. Перевести результати ЗНО в дихотомічну шкалу з прохідним балом 140.
6. За допомогою Q-критерію Кохрена перевірити, чи успішність складання ЗНО з різних предметів (набрав чи не набрав 140 балів) суттєво відрізняється.
7. Провести попарну перевірку успішності складання ЗНО (набрав чи не набрав 140 балів) за G-критерієм знаків.
8. Порівняти попарно середні результати ЗНО з різних третіх предметів з одного боку та української і 2-го предмета з іншого боку за критеріями t Стюдента для парних вибірок, T Вілкоксона та G знаків.
9. Провести попарну перевірку успішності складання ЗНО (набрав чи не набрав 140 балів) з різних третіх предметів з одного боку та з української й 2-го предмета з іншого боку за G -критерієм знаків.
10. Зробити загальні висновки.

Порядок виконання роботи № 4

1. Створити у своїй теці підтеку під назвою ...4

(замість трикрапки своє прізвище англійською).

2. Відкрити системний файл із попередньої практичної роботи та зберегти його у теці ...4 під назвою ...4 (замість трикрапки своє прізвище англійською).
3. Видаляємо змінні “Z-оцінки...”
4. Вибираємо в меню Аналіз — Непараметричні критерії — К пов’язаних вибірок. Тип тесту вибираємо тільки Фрідмана. Переносимо у вікно змінні Українська..., 2-й предмет та пЗ. Натискаємо Гаразд.
5. Вибираємо в меню Аналіз — Непараметричні критерії — 2 пов’язані вибірки. Тип тесту вибираємо Уїлкоксона та Знаків. Переносимо у вікно змінні Українська... та 2-й предмет. Натискаємо Гаразд.
6. Так само порівнюємо середні двох інших пар предметів: української з 3-м предметом та 2-го предмета з 3-м.
7. Вибираємо в меню Аналіз — Порівняти середні — Т перевірка парних вибірок. Переносимо у вікно змінні Українська... та 2-й предмет. Натискаємо Параметри. Довірчий інтервал встановлюємо 90 (тому що видається двобічний інтервал, а нас цікавить одnobічний 95%). Продовжити. Натискаємо Гаразд.

8. Так само порівнюємо середні двох інших пар предметів: української з 3-м предметом та 2-го предмета з 3-м.
9. Експортуємо файл Вивід у ...4.odt (замість трикрапки своє прізвище англійською) та закриваємо його.
10. Аналізуємо таблиці файлу...4.odt. Першим ми використовували χ^2 -критерій Фрідмана. Якщо Асимпт. знач. менша від 0,05, то пишемо під цією таблицею: “За критерієм Фрідмана більша кількість абітурієнтів отримують значно кращі бали ЗНО з одного предмета, ніж із іншого на рівні значущості (вказуємо Асимпт. знач.). Далі потрібно зробити попарну перевірку.”
11. Якщо Асимпт. знач. більша від 0,05, то пишемо: “За критерієм Фрідмана немає підстав стверджувати, що більша кількість абітурієнтів отримують значно кращі бали ЗНО з одного предмета, ніж із іншого.”
12. Вибираємо в таблиці предмети з найбільшим та найменшим середнім рангом. Під таблицями Ранги та Критерії перевірки з порівняння цих двох предметів у випадку, коли *Знач. (двобічна)* менша від 0,017 пишемо: ”За Т-критерієм Вілкоксона, частіше бали ЗНО з ... значно вищі, ніж із ... на рівні значущості ... (вказану двобічну значущість

множимо на 3).

13. Під таблицями Ранги та Критерії... двох інших пар предметів позитивний висновок пишемо у випадку, коли *Знач. (двобічна)* менша від 0,05. Цього разу двобічну значущість не потрібно множити на 3.
14. Якщо ж *Знач. (двобічна)* більша від 0,017 для пари з найбільшим та найменшим середніми рангами та від 0,05 для двох інших пар, то пишемо: “За Т-критерієм Вілкоксона, немає підстав стверджувати, що бали ЗНО з ... частіше значно вищі (чи нижчі), ніж з ... на рівні значущості (вказану двобічну значущість множимо на 3 для пари крайніх предметів).”
15. Аналогічно записуємо висновки під таблицями Частоти та Критерії перевірки чи вищі бали ЗНО з ... є в більшій кількості абітурієнтів, ніж із ... та під таблицями Перевірка парних вибірок чи бали ЗНО з ... в середньому вищі, ніж із ... на рівні значущості ...
16. Якщо за t-критерієм Стьюдента для парних вибірок середнє значення балів ЗНО з ... є значно вище, ніж із ..., то, крім цього висновку дописуємо також “Різниця з надійністю 0,95 більша від ... балів.” (Вибираємо ближче до 0 значення із стовпчиків Нижня, Верхня. Інакше пишемо: “Немає підстав стверджувати, що середні значення балів ЗНО з ... та

... різні на рівні значущості ...

17. Зберігаємо ...4.odt.
18. Для перекодування результатів ЗНО в дихотомічну шкалу вибираємо в меню Перетворення — Перекодувати до інших змінних. Переносимо змінну Українська..., 2-й предмет та Бали... стрілочкою. Натискаємо на укр і активізується вікно Нова змінна. У ній набираємо “укр140” (мітка — Чи набрав 140 балів з української) і натискаємо Зміна.
19. Натискаємо Попередні і нові значення, потім “Діапазон значень від найменшого до вказаного” і вводимо 140, а в полі Значення вводимо 0. Натискаємо Додати. Далі натискаємо кнопку Усі інші значення, а в полі Значення вводимо 1. Знову натискаємо Додати. Продовжити.
20. Те ж саме проробляємо зі змінними п2 та п3. Закінчивши цю роботу, натискаємо Гаразд. Зберігаємо системний файл.
21. Переходимо до вікна Перегляд змінних, відводимо ширину 1 і 1 стовпчик; Міра — Порядкова. Зберігаємо системний файл.
22. Вибираємо в меню Аналіз — Непараметричні критерії — К пов’язаних вибірок. Тип тесту вибираємо тільки Кохрена. Переносимо у вікно усі змінні Чи набрав ... Натискаємо Гаразд.
23. Вибираємо в меню Аналіз — Непараметричні критерії — 2 пов’язані вибірки. Тип тесту вибираємо тільки Знаків. Переносимо у вікно перші

дві змінні Чи набрав ... Натискаємо Гаразд.

24. Так само порівнюємо успішне складання ЗНО з двох інших пар предметів: української з 3-м предметом та 2-го предмета з 3-м.
25. Експортуємо файл Вивід у ...41.odt. Копіюємо з нього нову інформацію, переносимо в файл ...4.odt та зберігаємо цей файл. Закриваємо Вивід.
26. Записуємо у файлі ...4.odt висновки, чи є суттєва різниця між успішним складанням ЗНО з різних предметів, вказуючи значущість (у випадку Q-критерію Кохрена — асимптотичну, а у випадку G-критерію знаків — двобічну).
27. Зберігаємо файли.
28. Вибираємо в меню Дані — Поділити файл. Натискаємо Порівняти групи. Переносимо змінну Вибір... Натискаємо Гаразд.
29. Вибираємо в меню Аналіз — Порівняти середні — Т перевірка парних вибірок. Переносимо у вікно змінні Українська... та Бали... Натискаємо Параметри. Довірчий інтервал встановлюємо 90. Продовжити. Натискаємо Гаразд.
30. Так само порівнюємо середні значення 2-го та 3-го предметів.
31. Вибираємо в меню Аналіз — Непараметричні критерії — 2 пов'язані вибірки. Тип тесту вибираємо Уїлкоксона та Знаків. Переносимо у

- вікно змінні Українська... та п2. Натискаємо Гаразд.
32. Так само порівнюємо середні значення п2 та Бали...
33. Вибираємо в меню Аналіз — Непараметричні критерії — 2 пов'язані вибірки. Тип тесту вибираємо тільки Знаків. Переносимо у вікно перші 2 змінні Чи набрав з ... Натискаємо Гаразд.
34. Так само порівнюємо успішне складання ЗНО з двох інших пар предметів: української з 3-м предметом та 2-го предмета з 3-м.
35. Вибираємо в меню Аналіз — Непараметричні критерії — 2 пов'язані вибірки. Тип тесту вибираємо тільки Знаків. Переносимо у вікно змінні Чи набрав ... з української та Чи набрав ... з 3-го предмета Натискаємо Гаразд.
36. Вибираємо в меню Аналіз — Непараметричні критерії — 2 пов'язані вибірки. Тип тесту вибираємо тільки Знаків. Переносимо у вікно змінні Чи набрав ... з історії України та Чи набрав ... з 3-го предмета Натискаємо Гаразд.
37. Експортуємо файл Вивід у ...42.odt. Копіюємо з нього нову інформацію і переносимо в файл ...4.odt. Зберігаємо ...4.odt.
38. Записуємо висновки у файлі ...4.odt після кожної перевірки. Пам'ятаємо, що критерій Вілкоксона можна використовувати тільки при обсягах вибірки

$n > 14$.

39. Робимо загальні висновки. Зберігаємо ...4.odt.

40. Надсилаємо файли ...4.sav та ...4.odt на електронну пошту викладачу для перевірки.

§ 5. Виявлення відмінностей у рівнях

Мета роботи: Навчитися виявляти наявність різниці у рівнях досліджуваної ознаки у незалежних вибірках даних.

Вступна інформація. Виявляти наявність різниці у рівнях досліджуваної ознаки числової шкали потрібно за допомогою t-критерію Стьюдента для пари незв'язаних груп навіть якщо розподіл не є нормальним (але все ж неперервний).

В програмі PSPP при застосуванні t-критерію Стьюдента для пари незв'язаних груп автоматично перевіряється однорідність дисперсій за критерієм Левене. Цей критерій добре працює і тоді, коли розподіл ознаки не є нормальним. Правда, він не завжди є найпотужнішим.

Якщо ж дисперсії виявляться неоднорідними, то будемо користуватися варіантом t-критерію Стьюдента для пари незв'язаних груп, в якому не припускається рівність дисперсій. Для його використання бажано, щоб обсяги вибірок груп були однаковими або великими, особливо, якщо розподіл суттєво відхиляється від нормального.

За U-критерієм Манна Вітні можна перевірити, чи

відрізняються імовірності вищого значення ознаки у навмання вибраних осіб двох різних груп. Те ж саме перевіряється за Н-критерієм Крускала-Волліса у випадку 3-х і більше груп.

В програмі PSPP Н-критерій Крускала-Волліса реалізований тільки для великих обсягів вибірки. В такому разі у випадку двох груп він даватиме такий же результат, як і U-критерій Манна Вітні.

Щоб виявити, між якими саме змінними є значуща різниця, можна використати метод множинного порівняння Холма-Бонферроні. Якщо на більшому інтервалі порівняння різниця не виявлена, то на меншому уже не потрібно порівнювати.

Завдання роботи № 5:

1. Перевірити наявність різниці у рівнях результатів ЗНО по всіх предметах та середнього балу атестата залежно від наявності сільського коефіцієнта та вибору 3-го предмета за t-критерієм Стьюдента та U-критерієм Манна—Вітні.
2. Перевірити наявність різниці у рівнях результатів ЗНО по всіх предметах та середнього балу атестата залежно від суміщення змінних Сільський

коефіцієнт і Вибір 3-го предмета за t-критерієм Стьюдента та H-критерієм Крускала—Волліса.

3. Перевірити наявність різниці у рівнях результатів ЗНО по всіх 3-х предметах залежно від рівня середнього бала атестата (4 групи) за t-критерієм Стьюдента та H-критерієм Крускала—Волліса.
4. Зробити загальні висновки.

Порядок виконання роботи № 5

1. Створити на робочому столі теку під назвою ...5.
2. Відкрити системний файл із попередньої практичної роботи та зберегти його у теці ...5 під назвою ...5.
3. Вибираємо в меню Аналіз — Порівняти середні — Т-перевірка незалежних вибірок. У поле Змінна групування переносимо Вибір..., а у поле Змінні для перевірки — Українська..., п2, Бали... та Середній бал атестата. Натискаємо Визначити групи. Встановлюємо Значення групи 1: 1, а Значення групи 2: 0. Продовжити. Натискаємо Параметри. Довірчий інтервал змінюємо на 90%. Продовжити. Натискаємо Гарзд. Експортуємо Вивід у ...5.odt. Після цього закриваємо файл Вивід.
4. Якщо в 2-й таблиці у стовпчику “Перевірка Лівія рівності дисперсій” Значущість більша від 0,05, то немає підстав стверджувати, що групи неоднорідні за дисперсією. Тому звертаємо увагу на Значущість

(двобічна) у рядку “Припускаємо рівність дисперсій”. У цьому випадку у файлі ...5.odt робимо запис: “Немає підстав стверджувати, що групи неоднорідні за дисперсією за критерієм Левене на рині значущості ...”.

5. У протилежному випадку нам потрібна Значущість (двобічна) у рядку “Не припускаємо рівності дисперсій”, а у файлі ...5.odt записуємо: “групи неоднорідні за дисперсією за критерієм Левене на рині значущості ...”.
6. Якщо Значущість (двобічна) більша від 0,05, то за t -критерієм Стюдента немає підстав стверджувати, що є різниця середніх значень у групах. У протилежному випадку робимо висновок, на якому рівні значущості і в якій групі середній рівень ознаки вищий. Також дописуємо “З надійністю 0,95 ця різниця не менша від ... балів” (якщо межі надійних інтервалів мають однакові знаки) у файлі ...5.odt під 2-ю таблицею та зберігаємо цей файл.
7. Робимо перевірку за U -критерієм Манна—Вітні. Аналіз — Непараметричні критерії — К незалежних вибірок. Вибираємо тип тесту Н Крускала-Волліса. У поле Змінна групування переносимо Вибір..., а у поле Список змінних для тестування — усі 4 числові змінні. Натискаємо

Визначити групи. Встановлюємо Нижнє обмеження 0, Верхнє обмеження 1. Продовжити. Гаразд. Експортуємо файл Вивід у ...51.odt. Після цього закриваємо файл Вивід.

8. Переносимо інформацію з ...51.odt у ...5.odt. Зберігаємо файл ...5.odt. Записуємо в ньому висновки і знову зберігаємо. Якщо Асимпт. знач. (у випадку U-критерію Манна—Вітні) менша від 0,05, то записуємо, який 3-й предмет вибрали абітурієнти, які частіше мають кращі оцінки з відповідного предмета, але обсяги вибірок можуть бути недостатньо великими для застосування цього критерію. Інакше напишемо, що немає підстав стверджувати, що рівень успішності залежить від вибору 3-го предмета. Вказуємо рівень значущості.
9. Повторюємо цю процедуру зі змінною Сільський коефіцієнт. Експортуємо Вивід у ...52.odt. Після цього закриваємо файл Вивід.
10. Переносимо інформацію з ...52.odt у ...5.odt. Зберігаємо файл ...5.odt. Записуємо в ньому висновки і знову зберігаємо.
11. Утворимо суміщення змінних Вибір... та Сільський коефіцієнт. Для цього вибираємо Перетворення — Обчислити. У полі Змінна для обробки записуємо скв. У полі Числові вирази пишемо 10, знак

множення, переносимо змінну Вибір..., знак додавання, переносимо змінну ...коефіцієнт Натискаємо Гаразд. Зберігаємо системний файл.

12. У вікні Перегляд змінних встановлюємо для скв Ширина 2, Стовпчики 2, Знаків після коми 0, Міра Порядкова. Зберігаємо системний файл. Якщо якесь значення змінна набуває менше трьох раз, відмічаємо його, як пропущене.
13. Перевіряємо за t-критерієм Стюдента усі числові змінні за змінною скв. Значення групи 1: 1, Значення групи 2: 10 (якщо вони не пропущені). Потім те ж саме робимо із значеннями 1:0, 2:11 (навхрест, якщо вони не пропущені).
14. Експортуємо Вивід у ...53.odt. Після цього закриваємо файл Вивід.
15. Переносимо інформацію у ...5.odt. Зберігаємо файл ...5.odt. Записуємо в ньому висновки і знову зберігаємо.
16. Робимо також перевірку за H-критерієм Крускала-Волліса, чи імовірність отримати кращі результати залежить від суміщення змінних ск та в, встановивши нижнє обмеження 0, а верхнє 11. Підкреслюємо у висновку, що обсяги вибірок можуть бути недостатньо великими для застосування цього критерію. Попарну перевірку за U-критерієм Манна—Вітні не робимо, бо обсяги

вибірок малі.

17. Перевіряємо за t-критерієм Стюдента змінні Українська, п2, Бали... зі змінною Середній бал: 4 групи із кожною з 6 пар значень, починаючи від більших інтервалів порівняння. Якщо на більшому інтервалі порівняння різниця не виявлена, то на меншому уже не потрібно порівнювати.
18. Робимо також перевірку за H-критерієм Крускала-Волліса, чи імовірність отримати кращі результати ЗНО залежить від рівня середнього балу атестата, встановивши нижнє обмеження 1, а верхнє 4, тобто не попарно.
19. Знову пишемо висновки у файлі ...5.odt, перенісши у нього всі таблиці. Підкреслюємо у висновку, що обсяги вибірок можуть бути недостатньо великими для застосування H-критерію Крускала-Волліса.
20. Зберігаємо файли.
21. Робимо загальні висновки, в яких пояснюємо отримані результати. Зберігаємо ...5.odt.
22. Надсилаємо файли з розширеннями ...5.sav та ...5.odt на електронну пошту викладачу для перевірки.

§ 6. Перевірка однорідності двох вибірок

Мета роботи: Навчитися порівнювати між собою вибіркві розподіли, а також дисперсії.

Вступна інформація. У випадку нормального розподілу перевіряти гіпотезу про однорідність дисперсій можна не тільки за критерієм Левене, але й за іншим критеріями, зокрема F -критерієм Фішера, який часто виявляється потужнішим.

Порівнювати між собою розподіли можна, маючи дві вибірки великого обсягу. Але на даному занятті ми це будемо вчитися робити на прикладі вибірок середнього обсягу, бо лише такі дані ми маємо.

Найпростіше в PSPP перевіряти гіпотезу про однакові розподіли можна за критерієм серій Вальда-Вольфовиця, але він реалізований тільки для великих обсягів вибірки (більших від 20). Крім того програма видає двобічну значущість, тоді як насправді потрібна онобічна, яка отримується з двобічної діленням на 2.

Для застосування критерію серій Вальда-Вольфовиця слід упорядкувати дані за змінною, яка досліджується, а потім запустити на перевірку альтернативну змінну, яка ділить вибірку на дві групи, що порівнюються. При цьому порогове значення слід встановити Середнє значення, бо мода та переважно й медіана належатимуть до однієї групи, і розділення не відбудеться.

Коли розподіл на групи здійснюється за високими й низькими значеннями іншої змінної числової (або порядкової) шкали, альтернативну змінну можна не утворювати, а тільки задати порогове значення для

числової змінної.

Якщо кількість серій дорівнює критичному значенню (з таблиці) або менша від нього, то розподіли відрізняються між собою.

Таблиця критичних значень одnobічного критерію серій Вальда-Вольфовиця рівня значущості 0,05 ($n \geq m$)

2, якщо $m=2, n \geq 8$, або $m=3, 5 \leq n \leq 9$, або $m=4, 4 \leq n \leq 5$;

3, якщо $m=3, n \geq 10$, або $m=4, 6 \leq n \leq 11$, або $m=5, 5 \leq n \leq 8$, або $m=6, n=6$;

4, якщо $m=4, n \geq 12$, або $m=5, 9 \leq n \leq 13$, або $m=6, 7 \leq n \leq 9$, або $m=7, 7 \leq n \leq 8$;

5, якщо $m=5, n \geq 14$, або $m=6, 10 \leq n \leq 14$, або $m=7, 9 \leq n \leq 11$, або $m=8, 8 \leq n \leq 9$;

6, якщо $m=6, n \geq 15$, або $m=7, 12 \leq n \leq 16$, або $m=8, 10 \leq n \leq 13$, або $m=9, 9 \leq n \leq 11$, або $m=10, n=10$;

7, якщо $m=7, n \geq 17$, або $m=8, 14 \leq n \leq 17$, або $m=9, 12 \leq n \leq 14$, або $m=10, 11 \leq n \leq 12$, або $m=11, n=11$;

8, якщо $m=8, n \geq 18$, або $m=9, 15 \leq n \leq 19$, або $m=10, 13 \leq n \leq 16$, або $m=11, 12 \leq n \leq 14$, або $m=12, n=12$;

9, якщо $m=9, n \geq 20$, або $m=10, n \geq 17$, або $m=11, 15 \leq n \leq 17$, або $m=12, 13 \leq n \leq 15$, або $m=13, 13 \leq n \leq 14$;

- 10, якщо $m=11$, $n \geq 18$, або $m=12$, $16 \leq n \leq 19$, або $m=13$, $15 \leq n \leq 17$, або $m=14$, $14 \leq n \leq 15$;
- 11, якщо $m=12$, $n \geq 20$, або $m=13$, $n \geq 18$, або $m=14$, $16 \leq n \leq 18$, або $m=15$, $15 \leq n \leq 17$, або $m=16$, $n=16$;
- 12, якщо $m=14$, $n \geq 19$, або $m=15$, $n \geq 18$, або $m=16$, $17 \leq n \leq 18$, або $m=17$, $n=17$;
- 13, якщо $m=16$, $n \geq 19$, або $m=17$, $n \geq 18$, або $m=18$, $n=18$;
- 14, якщо $m=18$, $n \geq 19$, або $m=19$, $n \geq 19$;
- 15, якщо $m=20$, $n \geq 20$.

Завдання роботи № 6:

1. Порівняти розподіли успішності по третіх предметах за критеріями Смірнова та серій Вальда-Вольфовиця.
2. Перевірити гіпотезу про рівність дисперсій цих змінних за F -критерієм Фішера. Порівняти його значущість із значущістю критерію Левене з попередньої роботи.
3. Зробити висновки.

Порядок виконання роботи № 6

1. Створити на робочому столі теку під назвою ...б.
2. Завантажити програму PSPP. Відкрити системний файл із попередньої практичної роботи та зберегти

його у теці ...6 під назвою ...6.sav.

3. Упорядкуємо дані за значеннями змінної Бали... Для цього вибираємо в меню Дані — Упорядкувати спостереження. Переносимо змінну Бали..., а після неї також Вибір... Порядок — за зростанням. Натискаємо Гаразд. Зберігаємо файл.
4. Тепер вибираємо Аналіз — Непараметричні критерії — Серій. Встановлюємо Порогове значення — Середнє значення. Переносимо змінну Вибір... Гаразд.
5. Експортуємо Вивід у ...6.odt. Закриваємо Вивід.
6. Якщо обсяги обох вибірок більші від 20 і *Асимпт. знач. (двобічна)* лежить у межах від 0,1 до 1, то пишемо: “За критерієм серій Вальда на рівні значущості 0,05 немає підстав відкидати гіпотезу про однакові розподіли результатів ЗНО з ... та ...”.
7. *Асимпт. знач. (двобічна)* не може бути більшою від 1.
8. Якщо обсяги обох вибірок більші від 20 і *Асимпт. знач. (двобічна)* менша від 0,1, то пишемо: “За критерієм серій Вальда на рівні значущості 0,05 гіпотезу про однакові розподіли результатів ЗНО з ... та ... слід відкинути.
9. Якщо обсяг хоча б однієї вибірки не більший від 20,

то порівнюємо кількість серій з критичними значеннями з таблиці та робимо відповідні висновки.

10. Для перевірки гіпотези про однаковість розподілів результатів ЗНО з третіх предметів за критерієм Смірнова знайдемо правобічні границі їх емпіричних функцій розподілу. Для цього вибираємо в меню Перетворення — Ранжування спостережень. Вибираємо змінну Бали... і переносимо її у вікно Змінні. У вікно Критерій переносимо змінну Вибір... Типи рангів — Дробовий ранг. Продовжити. Збіги — Найбільший. Продовжити. Натискаємо Гаразд. Отримали змінну RпЗ. Вона містить значення емпіричних функцій розподілу обох 3-х предметів. Зберігаємо файл.
11. Переходимо до вікна Перегляд змінних. Для RпЗ встановлюємо Ширина — 6, Decimal — 4, Столпчики — 4. Знову зберігаємо файл.
12. Для порівняння значень емпіричних функцій розподілу обох 3-х предметів вводимо допоміжну змінну — Інша такого ж формату, як і RпЗ. Зберігаємо файл.
13. У вікні Перегляд даних набираємо значення змінної Інша. Верхнє значення RпЗ — це перший стрибок однієї з емпіричних функцій розподілу. Друга ж

поки-що не змінилася. Тому вводимо значення змінної Інша— 0.

14. Якщо у наступному рядку значення змінної Вибір... не змінюється у порівнянні з попереднім, то це означає, що змінюється та ж сама емп.ф.р. Інша залишається незмінною. Тому наступне значення змінної Інша копіюємо з попереднього її значення. Але якщо значення змінної Вибір... змінюється, то це означає, що в РпЗ записане нове значення тієї емп.ф.р., яка була в стовпчику Інша. Отже, ролі стовпчиків міняються місцями. Тому наступне значення змінної Інша копіюємо з попереднього значення змінної РпЗ. Але якщо при зміні значення змінної Вибір змінилося з 0 на 1, а пЗ при цьому — ні, то нове значення змінної РпЗ потрібно вставити в усі попередні клітинки змінної Інша з такими самими значеннями змінної пЗ.
15. Закінчивши вводити усі значення змінної Інша, зберігаємо системний файл.
16. Тепер порівняємо значення емпіричних функцій розподілу. Для цього знаходимо Перетворення — Обчислити. Змінна для обробки вводимо д. У віконечко Числові вирази переносимо знизу функція Abs. В дужки переносимо зліва РпЗ, ставимо знак —, а тоді переносимо Інша. Натискаємо Гаразд. Зберігаємо файл.

17. Формат змінної д встановлюємо такий же, як і Rp3. Зберігаємо системний файл.
18. Нам потрібно знайти найбільше значення змінної Інша. Щоб не знаходити його візуально, вибираємо Аналіз — Описова статистика — Частоти. У віконечко Змінні переносимо д. Статистика — вибираємо тільки Максимум. Натискаємо Гарзд. У файлі Вивід у нижній таблиці знаходимо цей Максимум.
19. Обсяги двох вибірок m, n записані у файлі ...6.odt на початку в таблиці Перевірка серій.
20. Обчислимо статистику Смірнова з поправкою Лемешка

$$S_{CM} = \left(D_{n,m} + (n+m)/(4.6 \cdot n \cdot m) \right) \cdot \sqrt{n \cdot m / (n+m)} .$$

Для цього вибираємо Перетворення — Обчислити. Набираємо Змінну для обробки КС. Відкриваємо дужку, пишемо Максимум змінної д (крапка замість коми) + в дужках сума обсягів вибірок, поділити на (в дужках) добуток обсягів вибірок на 4.6. Закриваємо дужку, ставимо знак множення. Переносимо знизу функцію SQRT. В дужках пишемо добуток обсягів вибірок, поділений на їхню суму. Натискаємо Гарзд. Зберігаємо файл. Експортуємо Вивід у ...61.odt. Закриваємо Вивід. Переносимо інформацію з ...61.odt у ...6.odt.

21. Потрібно порівняти її з 95% квантилем розподілу Колмогорова 1,358. (Квантиль рівня α можна наближено обчислити за формулою $\sqrt{\frac{1}{2} \ln \frac{2}{1-\alpha}}$.)
- Якщо отримана статистика КС більша від 1,358, то гіпотеза про рівність розподілів відкидається. Інакше немає підстав стверджувати, що розподіли — різні. Записуємо висновок у файлі ...6.odt. Зберігаємо цей файл.
22. Перевіримо рівність дисперсій за критерієм Фішера. Для цього вибираємо Аналіз — Описова статистика — Дослідити. Список залежних — Бали...; Список факторів — Вибір... Натискаємо Статистика — Описова статистика. Продовжити. Натискаємо Гаразд. Експортуємо файл Вивід у ...62.odt. Зберігаємо системний файл. Переносимо інформацію з ...62.odt у ...6.odt.
23. F-статистика дорівнює відношенню більшого Розсіювання з останньої таблиці Описова статистика до меншого. Записуємо це відношення у файлі ...6.odt. Подивімося, які обсяги вибірок відповідають чисельнику та знаменнику. Віднявши від них по 1, отримаємо кількості ступенів вільності чисельника та знаменника F-розподілу Беренса-Фішера-Снедекора. Записуємо ці кількості ступенів вільності у файлі ...6.odt.

24. Перетворення — Обчислити. Змінна для обробки — Знач. Її мітка — Точна значущість. Числові вирази: 2, знак мінус, 2, помножити, переносимо знизу функцію CDF.F. Її аргументи: відношення більшого Розсіювання до меншого, кількості ступенів вільності чисельника та знаменника відповідно. Натискаємо Гаразд. Зберігаємо файл.
25. Якщо $\text{Знач} > 0,05$, то немає підстав відхиляти гіпотезу про рівність дисперсій результатів ЗНО з 3-х предметів. Інакше пишемо, що гіпотеза про рівність дисперсій результатів ЗНО з 3-х предметів відхиляється на рівні значущості (пишемо значення змінної Знач). Записуємо висновок у файлі ...6.odt та зберігаємо.
26. Гіпотезу про рівність дисперсій за критерієм Левене ми перевіряли на попередньому занятті. Але щоб не шукати результати, зробимо цю перевірку ще раз іншим способом. Аналіз — Порівняти середні — Однофакторний ANOVA. Залежні змінні — переносимо п3. Фактор — Вибір... Статистика — Гомогенність. Гаразд. Експортуємо файл вивід у ...63.odt. Переносимо з нього інформацію у ...6.odt.
27. У таблиці Тест однорідності дисперсій (а не ANOVA) дивимось Знач. Якщо вона більша від 0,05, то немає підстав відкидати гіпотезу про рівність дисперсій за критерієм Левене. Інакше ця гіпотеза

відкидається на рівні Знач. Записуємо висновок у файл ...6.odt та зберігаємо.

28. Робимо загальні висновки. Зберігаємо ...6.odt.

29. Надсилаємо файли з розширеннями ...6.sav та ...6.odt на електронну пошту викладачу для перевірки.

§ 7. Виявлення різниці між імовірностями

Мета роботи: Навчитися перевіряти наявність впливу фактора на результативну ознаку, якщо вони задані у дихотомічній шкалі; досліджувати можливість застосування тих чи інших критеріїв для такої перевірки; робити обґрунтовані висновки.

Вступна інформація. Перевірка відсутності зв'язку між альтернативними змінними рівносильна перевірці гіпотези про рівність імовірностей двох випадкових подій. Пропонований у багатьох підручниках критерій перевірки останньої рівносильний χ^2 -критерію Пірсона. Проте, якщо хоча б одна теоретична частота менша від 10, рекомендується використовувати χ^2 -критерій Пірсона з поправкою на неперервність (Єйтса) або з поправкою найбільшої правдоподібності. Якщо ж є теоретична частота, менша від 5, то χ^2 -критерій Пірсона взагалі не можна використовувати. У цьому випадку залишається тільки точний критерій Фішера.

Точний критерій Фішера використовується також у випадку однобічної альтернативної гіпотези, наприклад,

якщо ми порівнюємо результати опитування після проведення певної роботи із результатами опитування іншої групи респондентів без неї і очікуємо покращення результатів, за умови, що відносна частота певних відповідей зростає. Для цього в таблиці подається Точна знач. (1-бічна).

Якщо ж фактор заданий у числовій або порядковій шкалі, то перевести його в дихотомічну можна, відділивши респондентів зі стенайнами, більшими від 5 від респондентів зі стенайнами, меншими від 5. При цьому близько 20% інформації втрачається, бо приблизно стільки респондентів мають стенайн 5. Тому наявність зв'язку між змінними числової шкали краще перевіряти іншими методами, які ми розглядатимемо пізніше.

Завдання роботи № 7:

1. Перевірити наявність зв'язку між ознаками “Вибір...” та “Сільський коефіцієнт”.
2. Перетворити змінну Середній бал на альтернативну, виключивши дані зі стенайном 5.
3. Перевірити наявність впливу ознак “Вибір...” і “Сільський коефіцієнт” на всі інші альтернативні змінні.
4. Перевірити наявність впливу ознаки “Середній бал атестата” на результати ЗНО по кожному предмету.
5. Зробити загальні висновки.

Порядок виконання роботи № 7

1. Створити на робочому столі теку під назвою ...7.
2. Відкрити системний файл із попередньої практичної роботи та зберегти його у теці ...7 під назвою ...7.sav.
3. Перевіримо, чи залежить імовірність вибору 3-го предмета для ЗНО від наявності сільського коефіцієнта, тобто чи існує зв'язок між цими ознаками. Для цього вибираємо в меню Аналіз — Описова статистика — Таблиці спряженості. Вибираємо змінну Сільський коефіцієнт і переносимо її у вікно Рядки. Змінну Вибір... переносимо у вікно Стопчики. Вибираємо статистику χ^2 . Продовжити. Відмічаємо комірки — Кількість та Очікувані (теоретичні частоти). Продовжити. Натискаємо Гаразд.
4. Експортуємо файл Вивід у файл із розширенням ...7.odt у своїй теці під назвою теки. Закриваємо Вивід.
5. Аналізуємо таблицю файлу Вивід. Якщо там є теоретичні частоти (знизу комірок) менші від 5, то критерій χ^2 не можна використовувати. Якщо там є теоретичні частоти (знизу комірок) менші від 10, то можна використовувати критерій χ^2 тільки з поправками на неперервність (Єйтса) або з

поправкою найбільшої правдоподібності. Записуємо це у файлі ...7.odt. Інакше звертаємо увагу на Асимпт. знач. Якщо якесь із них менше від 0,05, то пишемо, що ймовірності вибору 3-го предмета різні для осіб із сільським коефіцієнтом та тих, які його не мають за критерієм (-ями) Якщо ж усі Асимпт. знач. більші від 0,05, то пишемо, що за критеріями Пірсона, коефіцієнтом правдоподібності, Пірсона з поправкою на неперервність, немає підстав вважати, що ймовірність вибору 3-го предмета залежить від наявності сільського коефіцієнта.

6. У будь-якому випадку точний критерій Фішера можна використовувати, порівнявши Точна знач. (2-бічна) із 0,05. Якщо Точна знач. (2-бічна) $> 0,05$, то записуємо у файлі ...7.odt під відповідною таблицею: “За точним критерієм Фішера не можна стверджувати, що наявність сільського коефіцієнта впливає на ймовірність вибору 3-го предмета”.
7. Якщо Точна знач. (2-бічна) $< 0,05$, то записуємо у файлі ...7.odt під відповідною таблицею: “За точним критерієм Фішера на рівні значущості ... абітурієнти із сільським коефіцієнтом частіше вибирають ..., ніж ...”.
8. Зберігаємо ...7.odt.

9. Для перекодування змінної Середній бал атестата в дихотомічну шкалу спочатку упорядкуємо дані за її зростанням. Далі виберемо числа, які знаходяться між значеннями Середнього бала атестата зі стенойнами 4 і 5 та 5 і 6 відповідно. Напишемо їх у файлі ...7.odt під позначеннями Нсб та Всб.
10. Вибираємо в меню Перетворення — Перекодувати до інших змінних. Переносимо змінну Середній бал атестата стрілочкою. Натискаємо на сб і активізується вікно Нова змінна. У ній набираємо “сб2” (мітка — Середній бал атестата — 2 групи) і натискаємо Зміна.
11. Натискаємо Попередні і нові значення. Натискаємо кнопку “Діапазон значень від найменшого до вказаного” і вводимо Нсб, а в полі Значення вводимо 1. Натискаємо Додати. Далі натискаємо кнопку “Діапазон значень від вказаного до найбільшого ” і вводимо Всб, а в полі Значення вводимо 2. Натискаємо Додати. Далі — Продовжити — Гаразд.
12. Переходимо до вікна перегляд змінних і в змінній сб2 відводимо ширину 1 і 1 стовпчик; Міра — Порядкова. Зберігаємо .sav.
13. Вибираємо в меню Аналіз — Описова статистика — Таблиці спряженості. Змінні Сільський коефіцієнт та Вибір 3-го предмета переносимо у вікно Рядки. Сб2 та 3 змінні “Чи набрав...” переносимо у вікно

Стовпчики. Вибираємо статистику χ^2 .
Продовжити. Відмічаємо комірки — Кількість,
Рядок та Очікувані. Продовжити. Натискаємо
Гаразд.

14. Експортуємо файл Вивід у ...71.odt. Копіюємо з нього нову інформацію і переносимо в файл ...7.odt.
15. Записуємо у файлі ...7.odt висновки під кожною таблицею про можливість застосування критерію Пірсона та про наявність зв'язку між ознаками.
16. Зберігаємо ...7.odt.
17. Так само перевіряємо наявність впливу ознаки сб2 на укр140, п2140, п3140 В останньому випадку слід попередньо поділити файл на групи за вибором 3-го предмета.
18. Зберігаємо файли.
19. Записуємо висновки під кожною таблицею.
20. Робимо загальні висновки. Зберігаємо ...7.odt.
21. Надсилаємо файли з розширеннями ...7.sav та ...7.odt на електронну пошту викладачу для перевірки.

§ 8. Кореляційний аналіз

Мета роботи: Навчитися перевіряти значущість зв'язку між змінними та оцінювати його тісноту.

Вступна інформація. Якщо розподіли обох зв'язаних змінних — нормальні, то тісноту їх зв'язку

найкраще характеризує коефіцієнт кореляції Браве-Пірсона. Наявність зв'язку (значущість коефіцієнта кореляції) можна перевірити за допомогою t-розподілу Стюдента.

Якщо розподіл хоча б однієї з двох змінних відрізняється від нормального, то коефіцієнт кореляції Браве-Пірсона характеризує тільки тісноту їх лінійного зв'язку. У випадку, коли обсяг вибірки — великий, наявність зв'язку між змінними можна перевірити також за допомогою t-розподілу Стюдента.

Коефіцієнт кореляції Спірмена розраховується, по-суті, за тією ж формулою, що й коефіцієнт кореляції Браве-Пірсона, але не для самих змінних, а їх рангів.

Проблемою для обчислення коефіцієнта кореляції Спірмена становить наявність зв'язаних рангів. Поправки, які пропонуються у багатьох підручниках, є лише приблизними, а не точними. Але за допомогою програми PSPP коефіцієнт кореляції Спірмена можна обчислити точно, знайшовши ранги змінних та обчисливши для них коефіцієнт кореляції Браве-Пірсона.

У випадку декількох порядкових змінних обчислюється коефіцієнт конкордації W-Кендала. Його дуже зручно використовувати для перевірки ступеня узгодженості думок експертів, адже експерти найчастіше дають оцінки в порядковій шкалі.

Обчисливши коефіцієнт кореляції Браве-Пірсона

для альтернативних змінних, ми отримаємо коефіцієнт асоціації Пірсона. (У деяких підручниках він називається також ϕ -коефіцієнтом кореляції Гілфорда.) На минулому занятті ми, по-суті, перевіряли наявність зв'язку між альтернативними змінними. А коефіцієнт асоціації Пірсона ϕ оцінює тісноту цього зв'язку.

Якщо змінні належать до шкали назв, то перевірити наявність зв'язку між ними можна так само за допомогою χ^2 -критерію Пірсона. Але у цьому випадку обсяг вибірки має бути дуже великим, адже для його застосування теоретичні частоти можуть бути меншими від 5 (але не меншими від 2) не більше, ніж у 20% комірок, а точний критерій Фішера використовується лише для 4-польних таблиць. Тісноту зв'язку між змінними шкали назв характеризує V- коефіцієнт Крамера (хоча в літературі пропонуються також інші показники).

Нарешті, у випадку, коли одна зі змінних — альтернативна, а інша — числова, то коефіцієнт кореляції Браве-Пірсона між ними називається бісеріальним коефіцієнтом кореляції. Він називається також коефіцієнтом дискримінації та використовується для оцінки ефективності тестових завдань. На 5-му практичному занятті ми перевіряли наявність різниці між середніми значеннями двох числових змінних. По-суті, це була перевірка наявності впливу альтернативної ознаки, за якою здійснюється поділ на 2 групи, на середнє значення змінної. На цьому занятті ми оцінюватимо також тісноту

цього впливу.

Завдання роботи № 8:

1. Обчислити коефіцієнти кореляції Браве-Пірсона між усіма 4-ма числовими змінними попарно, перевірити їх значущість та оцінити тісноту лінійного зв'язку.
2. Обчислити коефіцієнти асоціації Пірсона між 6-ма альтернативними змінними попарно, перевірити їх значущість та оцінити тісноту зв'язку.
3. Обчислити бісеріальні коефіцієнти кореляції між 6-ма альтернативними та 4-ма числовими змінними попарно, перевірити їх значущість та оцінити тісноту зв'язку.
4. Обчислити коефіцієнти кореляції Спірмена між Середнім балом атестата та результатами ЗНО з 3-х предметів попарно, перевірити їх значущість, оцінити тісноту зв'язку та порівняти з коефіцієнтами кореляції Браве-Пірсона.
5. Обчислити V- коефіцієнт Крамера для змінних Середній бал атестата — 4 групи та Сільський коефіцієнт — 3-й предмет. Перевірити наявність та оцінити тісноту зв'язку.
6. Обчислити коефіцієнт конкордації W-Кендала для рангів з 3-х предметів ЗНО. Перевірити його значущість та оцінити тісноту зв'язку.

7. Обчислити коефіцієнти кореляції Браве-Пірсона та Спірмена між результатами ЗНО з третіх предметів окремо з іншими числовими змінними. Перевірити їх значущість та оцінити тісноту зв'язку.
8. Побудувати надійний інтервал для найбільш значущого коефіцієнта кореляції з надійністю 0,95.
9. Зробити загальні висновки.

Порядок виконання роботи № 8

1. Створити на робочому столі теку під назвою ...8.
2. Завантажити програму PSPP. Відкрити системний файл із попередньої практичної роботи та зберегти його у теці ...8 під назвою ...8.sav.
3. Вибираємо в меню Аналіз — Двовимірна кореляція. Перевірка значущості — Двобічна. Відмічаємо Позначити значущі кореляції. Переносимо у вікно спочатку всі 4 числові змінні, а потім 8 альтернативних змінних. Натискаємо Гаразд.
4. Експортуємо файл Вивід у ...8.html та закриваємо його.
5. Аналізуємо таблиці файлу...8.html. На перетині перших 4-х рядків та стовпчиків подаються значення коефіцієнтів кореляції Браве-Пірсона між змінними та їх значущості. Якщо *Знач. (двобічна)* більша від 0,05, то немає підстав вважати, що між

змінними є зв'язок. Інакше робимо висновок про тісноту лінійного зв'язку за значенням *Кореляція Пірсона*. (Якщо $r \geq 0,9$, то лінійний зв'язок дуже сильний; якщо $0,7 \leq r < 0,9$, то сильний; якщо $0,5 \leq r < 0,7$, то середній, якщо $0,3 \leq r < 0,5$, то помірний; якщо $0,2 \leq r < 0,3$, то слабкий, а якщо $r < 0,2$, то дуже слабкий.) Зауважуємо, між якими змінними лінійний зв'язок найсильніший, а між якими — найслабший (або найменш значущий) та записуємо відповідний висновок у текстовому файлі під заголовком “Оцінка тісноти зв'язку числових змінних”.

6. На перетині останніх 6-ти рядків та стовпчиків подаються значення коефіцієнтів асоціації Пірсона між альтернативними змінними та їх значущості. Аналізуємо їх та записуємо відповідні висновки під заголовком “Оцінка тісноти зв'язку альтернативних змінних”.
7. На перетині останніх 6-и рядків та перших 4-х стовпчиків подаються значення бісеріальних коефіцієнтів кореляції та їх значущості. Аналізуємо їх та записуємо відповідні висновки. Зрозуміло, що найтісніший зв'язок буде між числовою змінною та альтернативною, утвореною з неї. Тому обмежуємося розглядом її кореляції з іншими альтернативними змінними. Зберігаємо ...8.txt.

8. Знайдемо ранги числових змінних. Для цього вибираємо в меню Перетворення — Ранжувати спостереження. Переносимо числові змінні. Натискаємо Типи рангів. Вибираємо Ранг. Продовжити. Натискаємо Збіги. Вибираємо Середній. Продовжити. Натискаємо Гаразд. Зберігаємо системний файл.
9. Вибираємо в меню Аналіз — Двовимірна кореляція. Перевірка значущості — Двобічна. Відмічаємо Позначити значущі кореляції. Переносимо у вікно щойно отримані 4 рангові змінні. Натискаємо Гаразд.
10. Експортуємо файл Вивід у ...8.odt. Записуємо у ньому висновки щодо коефіцієнтів кореляції Спірмена та порівнюємо їх із коефіцієнтами кореляції Браве-Пірсона. Також переносимо у цей файл висновки з ...8.txt. Зберігаємо ...8.odt.
11. Вибираємо в меню Аналіз — Непараметричні критерії — К пов'язаних вибірок. Тип тесту — W Кендала. Переносимо у вікно 3 рангові змінні результатів ЗНО з різних предметів. Натискаємо Гаразд.
12. Експортуємо файл Вивід у ...81.odt. Копіюємо з нього нову інформацію і переносимо в файл ...8.odt.
13. Якщо асимптотична значущість W-коефіцієнта

конкордації Кендала менша від 0,05, то записуємо висновок про узгодженість результатів ЗНО з різних предметів, а якщо менша, то про неузгодженість. Зберігаємо ...8.odt.

14. Проранжуємо результати ЗНО окремо для груп абітурієнтів, які вибрали різні 3-і предмети. Для цього Для цього вибираємо в меню Перетворення — Ранжувати спостереження. Переносимо числові змінні. Переносимо у поле Критерій змінну Вибір 3-го предмета. Натискаємо Типи рангів. Вибираємо Ранг. Продовжити. Натискаємо Збіги. Вибираємо Середній. Продовжити. Натискаємо Гарзд. Зберігаємо системний файл.
15. Вибираємо в меню Дані — Поділити файл. Вибираємо Порівняти групи. Переносимо 3-й предмет. Натискаємо Гарзд. Зберігаємо системний файл.
16. Вибираємо в меню Аналіз — Двовимірна кореляція. Перевірка значущості — Двобічна. Відмічаємо Позначити значущі кореляції. Переносимо у вікно щойно отримані 4 рангові змінні. Натискаємо Гарзд.
17. Вибираємо в меню Аналіз — Двовимірна кореляція. Перевірка значущості — Двобічна. Відмічаємо Позначити значущі кореляції. Переносимо у вікно 4

числові змінні. Натискаємо Гаразд.

18. Експортуємо файл Вивід у ...82.odt. Копіюємо з нього нову інформацію і переносимо в файл ...8.odt.
19. Записуємо у файлі ...8.odt висновки щодо коефіцієнтів кореляції Спірмена та Браве-Пірсона результатів ЗНО 3-х предметів з іншими змінними. Зберігаємо ...8.odt.
20. Вибираємо серед усіх знайдених коефіцієнтів кореляції найбільш значущий. Для знаходження для нього надійного інтервалу заходимо в програму MS EXCEL. Знаходимо функцію FISHER (Z-перетворення Фішера) від цього коефіцієнта кореляції.
21. У 2-ій комірці вводимо 1,96 (квантиль рівня 0,975 Z-розподілу), а в 3-й знаходимо SQRT (квадратний корінь) з N-3, де N — обсяг відповідної вибірки. Після цього в 4-й комірці ділимо 1,96 на цей корінь. Отримали половину надійного інтервалу.
22. Віднімаємо та додаємо до перетворення Фішера половину довжини надійного інтервалу. Результати зберігаємо в інших двох комірках.
23. Знаходимо FISHERINV від цих двох чисел. Це й будуть межі надійного інтервалу для відповідного коефіцієнта кореляції надійності 0,95. Записуємо цей інтервал у ...8.odt. Зберігаємо файли.

24. Робимо загальні висновки. Зберігаємо ...8.odt.

25. Надсилаємо усі отримані файли на електронну пошту викладачу для перевірки.

§ 9. Дисперсійний аналіз.

Мета роботи: Навчитися перевіряти значущість різниці середніх значень ознаки у різних групах за допомогою дисперсійного аналізу та методу множинних порівнянь, оцінювати ступінь впливу факторів на залежну змінну.

Вступна інформація. Хоча дисперсійний аналіз розроблявся для вивчення нормально розподіленої залежної змінної, згодом виявилось, що його можна успішно застосовувати і тоді, коли розподіл залежної змінної відрізняється від нормального.

Якщо обсяги груп приблизно однакові, то дисперсійний аналіз можна застосовувати і тоді, коли дисперсії залежної змінної у групах суттєво відрізняються.

Програма PSPP дає можливість використовувати для незв'язаних вибірок однофакторний дисперсійний аналіз із перевіркою контрастів, а також мультифакторний дисперсійний аналіз.

Завдання роботи № 9:

1. Перевірити за допомогою двохфакторного дисперсійного аналізу залежність середнього бала атестата та результатів ЗНО з 3-х предметів від наявності сільського коефіцієнта, вибору третього предмета та взаємодії цих факторів.
2. Перевірити за допомогою однофакторного дисперсійного аналізу залежність середніх балів ЗНО від рівнів середнього бала атестата. Виявити, між якими саме групами є розбіжності. Оцінити тісноту зв'язку.
3. Перевірити залежність середнього бала атестата та результатів ЗНО з 2-х предметів від результатів ЗНО з предмета, найбільше пов'язаного з середнім балом атестата. Виявити, між якими саме групами є розбіжності. Оцінити тісноту зв'язку.
4. Зробити загальні висновки.

Порядок виконання роботи № 9

1. Створити на робочому столі теку під назвою ...9.
2. Відкрити системний файл із попередньої практичної роботи та зберегти його у теці ...9 під назвою ...9.sav.
3. Вибираємо в меню Аналіз — Одновимірний аналіз. Переносимо фіксовані фактори: Сільський (галузовий) коефіцієнт та Вибір 3-го предмета.

Залежна змінна — Середній бал атестата.
Натискаємо Гаразд. Потім так само дослідимо результати ЗНО з усіх трьох предметів.

4. Експортуємо файл Вивід у ...9.odt та закриваємо його.
5. Порівнюючи значущість з 0,05, робимо висновок, чи залежить середній бал атестата та результати ЗНО від наявності сільського коефіцієнта, вибору 3-го предмета та взаємодії цих факторів.
6. Вибираємо в меню Аналіз — Порівняти середні — Однофакторний ANOVA. Фактор: Середній бал 4 групи. Залежні змінні: Українська, 2-й предмет, пЗ. Позначаємо Статистиска: Описова статистика та Гомогенність. Далі вибираємо Контрасти.
7. Набираємо коефіцієнт -1. Натискаємо Додати. Набираємо 0. Додати. Знову 0. Додати. І нарешті 1. Додати. Натискаємо стрілку Вперед.
8. Потім так само вводимо 0, Додати, -1, Додати, 0, Додати, 1, Додати. Вперед.
9. Після цього вводимо 3-й контраст: -1, 0, 1, 0. Вперед.
10. 4-й контраст. 0, 0, -1, 1. Вперед.
11. 5-й контраст 0, -1, 1, 0. Вперед.

12. 6-й контраст -1, 1, 0, 0. Але цього разу вже натиснемо Продовжити, а не Вперед, бо цей контраст — останній. Після Продовжити натискаємо Гаразд.
13. Експортуємо файл Вивід у ...91.odt та закриваємо його. Переносимо інформацію з нього в ...9.odt.
14. , В таблиці “Описова статистика” наводяться середні значення результатів ЗНО в різних групах. Упорядкувавши їх у порядку зростання, знайдемо інтервали порівняння між парами груп: між найбільшою та найменшою інтервал порівняння становить 4, між сусідніми 2, а два інші 3. Записуємо значення інтервалів порівняння між усіма парами груп для результатів ЗНО з кожного предмета.
15. “Тест однорідності дисперсій” звертаємо увагу на значущості в останньому стовпчику. Якщо вони більші від 0,05, то немає підстав відкидати гіпотезу про рівність дисперсій у групах. Записуємо відповідний висновок під цією таблицею.
16. Переходимо до аналізу таблиці ANOVA. Якщо Значущість більша від 0,05, то немає підстав стверджувати, що середнє значенням ЗНО з відповідного предмета залежить від рівня середнього бала атестата.

17. Якщо Значущість менша від 0,05, то записуємо:
“Середнє значення ЗНО з ... залежить від рівня середнього бала атестата на рівні значущості ...”.
Поділивши Суму квадратів Між групами на Загальну Суму квадратів, знайдемо, яка частка розбіжностей у результатах даного ЗНО пояснюється рівнем середнього бала атестата. Записуємо це. Добувши квадратний корінь із цього числа, знайдемо кореляційне відношення. Записуємо кореляційне відношення і оцінюємо силу кореляційного зв’язку. (Якщо кореляційне відношення $\geq 0,9$, то кореляційний зв’язок дуже сильний; якщо від 0,7 до 0,9, то сильний; якщо від 0,5 до 0,7, то середній, якщо від 0,3 до 0,5, то помірний; якщо від 0,2 до 0,3, то слабкий, а якщо кореляційне відношення $< 0,2$, то дуже слабкий.)
18. Аналізуємо таблицю Тести контрастності. У ній за t-критерієм Стюдента перевіряється, між якими групами (див. 5-ту роботу) є суттєва різниця в результатах ЗНО. Якщо в контрасті інтервал порівняння між групами становить 4, то відповідну значущість потрібно буде множити на 6, а якщо 3, то на 3. Якщо ж інтервал порівняння 2, то значущість не слід множити.
19. Записуємо, між якими групами є значуща різниця в результатах ЗНО та вказуємо рівень значущості.

20. Вибираємо змінну, яка найбільше пов'язана з середнім балом атестата. Вибираємо в меню Аналіз — Порівняти середні — Однофакторний ANOVA. Фактор: [вибрана змінна]—4 групи. Залежні змінні: Середній бал атестата та дві інші змінні. Позначаємо Статистиска: Описова статистика та Гомогенність. Далі вибираємо Контрасти (так само, як і раніше).
21. Експортуємо файл Вивід у ...92.odt та закриваємо його. Переносимо інформацію з нього в ...9.odt. Записуємо висновки під кожною таблицею.
22. Робимо загальні висновки. Зберігаємо ...9.odt.
23. Надсилаємо отримані файли на електронну пошту викладачу для перевірки.

§ 10 Регресійний аналіз

Мета роботи: Навчитися виявляти фактори, які впливають на результативну ознаку, будувати моделі залежностей та оцінювати їх якість і статистичну значущість.

Вступна інформація.

В програмі PSPP реалізована множинна лінійна та бінарна логістична регресії. Обчислюється коефіцієнт детермінації та значущість регресії в цілому та її коефіцієнтів. Будуються також надійні інтервали для

коефіцієнтів регресії. Крім того можна вивести прогнозовані значення та побачити, наскільки вони відрізняються від експериментальних даних.

Завдання роботи № 10:

1. Побудувати лінійні моделі із значущими коефіцієнтами середнього бала атестата та результатів ЗНО. Вказати коефіцієнти детермінації.
2. Вивести поля кореляції, щоб перевірити, чи лінійна модель є доцільною.
3. Побудувати бінарні логістичні моделі наявності сільського коефіцієнта та вибору 3-го предмета.
4. Зробити загальні висновки.

Порядок виконання роботи № 10

1. Створити на робочому столі теку під назвою ...10.
2. Відкрити системний файл із попередньої практичної роботи та зберегти його у теці ...9 під назвою ...10.sav.
3. Вибираємо в меню Аналіз — Регресія — Лінійна. Переносимо в залежні змінні Середній бал атестата, а в незалежні — Сільський (галузовий) коефіцієнт, Вибір 3-го предмета, Українська, п2 та п3. В Статистика вибираємо всі опції, а в Зберегти — жодної. Натискаємо Гаразд.

4. Потім поступово видаляємо змінні із найбільшою значущістю в останній таблиці, слідкуючи за тим, щоб коефіцієнт детермінації R-квадрат різко не зменшився. Якщо при вилученні чергової змінної коефіцієнт детермінації різко зменшиться, то його потрібно повернути до моделі і вилучення змінних припинити. Після цього повторюємо аналіз з опціями у Зберегти — Передбачені значення та Лишки. У вікні Змінні можна спостерігати прогнозовані моделлю регресії значення та їх похибки.
5. Експортуємо файл Вивід у ...10.odt та закриваємо його.
6. Записуємо під останньою таблицею отримане рівняння регресії, його коефіцієнт детермінації, значущість та стандартну похибку моделі.
7. Тепер те саме повторюємо із змінною Українська... Серед незалежних змінних буде Середній бал атестата, Сільський (галузевий) коефіцієнт, Вибір 3-го предмета, п2 та п3. Експортуємо Вивід у ...101.odt, переносимо інформацію в ...10.odt та записуємо висновки.
8. Далі так само досліджуємо п2, а також п3.
9. Тепер вибираємо Графіки — Точкова діаграма. Вісь Y — Середній бал атестата. Вісь X — Українська...

Потім Вісь X — п2, потім — п3.

10. Тепер Вісь Y — Українська..., а Вісь X — Середній бал атестата, пізніше п2 і п3.
11. Далі Вісь Y — п2, а Вісь X — інші 3 змінні.
12. І нарешті з Вісь Y — п3 будуємо ще 3 діаграми.
13. Експортуємо Вивід у ...10.pdf. Візуально оцінюємо за діаграмами лінійні залежності.
14. Вибираємо в меню Аналіз — Регресія — Бінарна логістична. Переносимо в залежні змінні Вибір 3-го предмета, а в незалежні — Сільський (галузевий) коефіцієнт, Українська, п2 та п3. В Параметри вибираємо всі опції. Натискаємо Гаразд.
15. Потім поступово видаляємо змінні із найбільшою значущістю в останній таблиці, слідкуючи за тим, щоб коефіцієнт детермінації Кокса і Снела різко не зменшився. Якщо при вилученні чергової змінної коефіцієнт детермінації різко зменшиться, то його потрібно повернути до моделі і вилучення змінних припинити.
16. Переносимо інформацію з файлу Вивід у ...10.odt.
17. Записуємо під останньою таблицею “Ймовірність вибору ... $= 1/(1+\exp\{-...\})$ ”. Вказуємо коефіцієнт детермінації Кокса і Снела та загальний виправлений відсоток.

18. Те ж саме повторюємо зі змінною “Сільський (галузевий) коефіцієнт”.
19. Робимо загальні висновки. Зберігаємо ...10.odt.
20. Надсилаємо отримані файли на електронну пошту викладачу для перевірки.

Література

1. Дж. Гласс, Дж. Стенли. Математические методы в психологии и педагогике. М. : Прогресс. 1976. 496 с.
<https://www.psychology-online.net/articles/doc-2268.html>
2. Б. Кожух, В. Н. Колесников. Программа PSPP в педагогических исследованиях. Петрозаводск : Изд-во ПерпГУ, 2017. 151 с.
https://edu.petrso.ru/files/upload/11199_1528976883.pdf
3. Високів І. Е. Математичні методи в психології.
https://stud.com.ua/50286/psihologiya/matematichni_metodi_v_psihologiyi
4. Климчук В. Математичні методи у психології : навчальний посібник.
https://www.researchgate.net/publication/281088252_Matematicni_metodi_u_psihologii_Navcalnij_posibnik